Convergent analysis of cDNA and short oligomer microarrays, mouse null mutants and bioinformatics resources to study complex traits

I. Ponomarev*†, G. L. Schafer†, Y. A. Blednov†, R. W. Williams‡, V. R. Iyer† and R. A. Harris†

†Waggoner Center for Alcohol and Addiction Research, The University of Texas at Austin, Austin, TX, ‡Department of Anatomy and Neurobiology, The University of Tennessee Health Science Center, Memphis, TN, USA

*Corresponding Author: Igor Ponomarev, Waggoner Center for Alcohol and Addiction Research, The University of Texas at Austin, 1 University Station A4800, Austin, TX 78712, USA. E-mail: piatut@mail.utexas.edu

Gene expression data sets have recently been exploited to study genetic factors that modulate complex traits. However, it has been challenging to establish a direct link between variation in patterns of gene expression and variation in higher order traits such as neuropharmacological responses and patterns of behavior. Here we illustrate an approach that combines gene expression data with new bioinformatics resources to discover genes that potentially modulate behavior. We have exploited three complementary genetic models to obtain convergent evidence that differential expression of a subset of genes and molecular pathways influences ethanol-induced conditioned taste aversion (CTA). As a first step, cDNA microarrays were used to compare gene expression profiles of two null mutant mouse lines with difference in ethanol-induced aversion. Mice lacking a functional copy of G protein-gated potassium channel subunit 2 (Girk2) show a decrease in the aversive effects of ethanol, whereas preproenkephalin (Penk) null mutant mice show the opposite response. We hypothesize that these behavioral differences are generated in part by alterations in expression downstream of the null alleles. We then exploited the WebQTL databases to examine the genetic covariance between mRNA expression levels and measurements of ethanol-induced CTA in BXD recombinant inbred (RI) strains. Finally, we identified a subset of genes and functional groups associated with ethanol-induced CTA in both null mutant lines and BXD RI strains. Collectively, these approaches highlight the phosphatidylinositol signaling pathway and identify several genes including protein kinase C beta isoform and preproenkephalin in regulation of ethanol-induced conditioned taste aversion. Our results point to the increasing potential of the convergent approach and biological databases to investigate genetic mechanisms of complex traits.

Keywords: Alcohol, brain, gene expression, in silico, knock-out

Received 25 February 2004, revised 3 May 2004, accepted for publication 11 May 2004

Oligonucleotide and cDNA microarrays have been employed recently to investigate genetic mechanisms of complex traits (Chesler et al. 2003; Hassan et al. 2003; Hitzemann et al. 2003; Schadt et al. 2003; Tabakoff et al. 2003; Wang et al. 2003). Various genetic mouse models including inbred strains (Chesler et al. 2003; Hitzemann et al. 2003), selected lines and null mutant (knockout, KO) mice (Hassan et al. 2003; Tabakoff et al. 2003; Tudor et al. 2002) have been favorite targets for such experiments. However, without additional information, it is virtually impossible to establish a direct functional link between transcriptional and behavioral variations. Choosing the most promising genes from large sets of significantly regulated candidates presents another challenging task. One initial solution is to exploit complementary techniques and resources that collectively highlight the possible involvement of specific genes in regulation of behavior. Here we used three independent genetic models to identify high priority candidate genes and functional groups potentially involved in regulation of ethanol-induced conditioned taste aversion (CTA). Two independently created KO mouse lines and a set of BXD recombinant inbred (RI) strains were investigated.

Hundreds of KO and transgenic mouse lines have been created to study gene function at cellular and behavioral levels. The range of phenotypes observed in these genetically modified mice most likely results from a combination of direct effects of the mutation and indirect changes that occur as compensation. The introduction of variant alleles of neighboring genes that ‘hitch-hike’ in linkage with the target gene is another potential source of phenotypic alterations (Gerlai 1996). The deletion of a gene can directly or indirectly result in changes in expression of other genes, which in turn can underlie behavioral changes observed in mutant animals. Even the inactivation of apparently unrelated genes can produce common behavioral alterations, perhaps due to downstream effects on gene expression. One approach to the
study of genetic mechanisms of behavior is to investigate transcript expression patterns in several genetically altered mouse models with common behavioral phenotypes. The goal of such studies is to identify common patterns of gene expression changes that are associated with specific phenotypic variants produced by inactivation of different genes. In the present study we used this approach to investigate expression profiles of two KO mouse lines that differed in expression of ethanol-induced CTA.

These lines of mice represent two important classes of molecules in neuronal signaling: G-protein-coupled inwardly rectifying potassium channel subunit 2 (GIRK2) and enkephalins. GIRK2-containing channels are the predominant GIRK channels in the brain and are coupled to several types of G-protein coupled receptors, including GABA\textsubscript{γ}, dopamine 2-like and opioid receptors. Deletion of this subunit also leads to a decrease in GIRK1 subunit expression, suggesting that the predominant channel isoform in the brain is dramatically decreased (Signorini et al. 1997). Preproenkephalin (PENK) is a peptide precursor of enkephalins – endogenous agonists of μ and δ opioid receptors. These gene deletions produce viable and relatively normal mutants (Konig et al. 1996; Signorini et al. 1997).

CTA is a learning-dependent behavioral change that is used to study hedonic effects of different chemical compounds (Risinger & Boyce 2002; Risinger & Cunningham 1998). CTA is usually seen as a reduction of consumption of some solution – for example, saccharin – after pairing with injections of a chemical with putative aversive effects. The magnitude of such a reduction is interpreted as the magnitude of CTA. It is believed that ethanol-induced CTA reflects ethanol’s motivational effects and negatively contributes to ethanol consumption. Evidence suggests that GIRK2 KO mice are less sensitive to the aversive effects of ethanol. When tested in a CTA paradigm, the GIRK2 mutant animals consumed more saccharin solution that has been paired with an aversive injection than control mice did (Hill et al. 2003). We have recently tested Penk null mutant mice on a number of ethanol phenotypes (Y.A. Blednov, manuscript in preparation). In contrast to GIRK2 KO mice, Penk-deficient animals showed more ethanol-induced aversion in a CTA paradigm. These studies identified ethanol-induced CTA as the only alcohol trait differentially regulated in both KO lines thus far, when measured under similar conditions.

In the present study we used cDNA microarrays to identify gene expression changes that occur as a result of the null mutation in GIRK2 and Penk lines. In addition we employed a complementary resource called WebQTL to detect large sets of mRNA transcripts that are genetically correlated with published measures of ethanol-induced CTA in BXD RI strains. An over-representation analysis was used to identify functional groups that were activated in both KO lines and in the BXD RI set. The goal was to select a subset of biologically relevant candidate genes and pathways potentially involved in regulation of ethanol-induced conditioned taste aversion.

Materials and Methods

Animals
GIRK2 KO mice (C57BL/6J \times 129/SvJ background) were generated as previously described (Signorini et al. 1997). Homozygous null mice were originally obtained from the colony maintained by Dr M. Stoffel and were later interbred with wild type mice with similar mixed genetic background to produce heterozygous mice (for details see Blednov et al. 2003). Heterozygous mating was maintained at the local colony to generate mutant and wild type littermates that were used as controls for the GIRK2 mutants. Penk null mutant mice were generated as previously described (Konig et al. 1996) and were later fully backcrossed (10 generations) to C57BL/6J genetic background (http://jaxmice.jax.org). The Penk null mutants B6.129-Penk-rt2m1Pig/J were obtained from the Jackson Laboratory (Bar Harbor, ME), where they are maintained by homozygous breeding. The Jackson Laboratory web site (above) recommends using C57BL/6J inbred mice as controls for these mutants. Therefore, C57BL/6J mice were obtained from the Jackson Laboratory and used as control animals in the present study. The tissue used in microarray experiments of these KO lines was dissected from the whole brain and comprised of a block of tissue roughly corresponding to −2 to −4 mm bregma (Paxinos & Franklin 2001) with colliculi, ventral hypothalamic and cortical structures removed. As such, the tissue block would contain ventral midbrain, caudal thalamic nuclei and dorsocaudal hypothalamic nuclei. This brain region was chosen because it contains numerous cell populations and nuclei that have been implicated in alcohol actions. Mice were 2.5–11 month old males. Eight GIRK2 KO, six GIRK2 WT, seven Penk KO and five Penk WT mice were used for this study. For hybridizations, samples were paired so that the ages of wild type and knockouts were matched.

Microarray procedures

cDNA microarray construction
cDNA clones from the mouse NIH BMAP (~11000 clones) and sequence verified IMAGE (~5000 clones) mouse clone sets (ResGen/Invitrogen, Carlsbad, CA), amplified and re-plated to 384-well plates. High-density arrays of cDNAs were printed on 1 \times 3-inch poly L-lysine coated slides at the University of Texas microarray facility and postprocessed with diversity of Texas microarray facility and postprocessed with succinic anhydride before use. cDNA synthesis and hybridization were performed based on published protocols from the Brown lab (http://brownlab.stanford.edu/protocols.html) with slight modifications. We have sequenced a subset of the BMAP and IMAGE clones to estimate the identification error rate. Based on this sequencing, the error rate was around 5%.

cDNA synthesis
Total RNA was isolated from tissue using TRIzol reagent (Invitrogen), resuspended in DEPC-H\textsubscript{2}O and cleaned using RNeasy (Qiagen, Valencia, CA). Total RNA 12.5–15\textmu g was
primed with 5 μg of an anchored T-mer [5′-T(20) VN-3′]. cDNA synthesis incorporated an aminoallyl-modified dUTP in a 2:3 ratio with dTTP (500 μM final concentration of dNTPs). The reaction was started by the addition of 400 U Superscript II Rnase H− reverse transcriptase (Invitrogen) to preheated (42 °C) samples. Samples were subsequently heat denatured and the RNA hydrolyzed by the addition of 1 N NaOH/0.5 M EDTA for 15 min at 67 °C. Samples were neutralized with 1 M HEPES, pH 7.5. cDNA was concentrated in filtered dH2O with Microcon-30 filters (Millipore, Amicon, Bedford, MA). cDNAs from KO and WT mice were labeled with either cyanine-5 (Cy-5) or cyanine-3 (Cy-3) dyes (Amersham, Buckinghamshire, UK) for 1 h at room temperature in the dark. Unlabeled cyanine dyes were removed with Qiagen purification kit (Qiagen). For some hybridizations, the fluorescent label was reversed for each genotype to avoid potential dye bias.

Hybridization

One KO and one WT sample from individual mice labeled with different dyes were combined and concentrated in a Microcon-30. Because there was an uneven number of KO and WT samples, two WT samples in each KO line were used twice in pairs with individual KO samples. Using the WT replicates should not elevate Type I error because, for our statistical analysis, we used fewer degrees of freedom than those limited by the number of subjects. Totally 8 Girk2 and 7 Penk KO/WT pairs were available for hybridization. The hybridization solution contained 3.5X SSC, 0.25% SDS; polyadenosine (10 μg), mouse Cot-1 DNA (1 μg), and yeast tRNA (5 μg) were added to each hybridization for non-specific hybridization blocking. Hybridization solution was heated to 98 °C for 2 min and immediately applied to a prepared slide. Hybridization was performed for 14–16 h at 61 °C in the dark. Slides were washed twice in 0.6X SSC/0.03% SDS at room temperature and twice in 0.06X SSC at room temperature, followed by a single dH2O rinse. All washes were performed in opaque Coplin jars (VWR, West Chester, PA). Slides were dried by centrifugation and scanned immediately on a Genepix 4000B scanner (Axon Instruments, Inc., Union City, CA).

Data filtering and analysis

Data were initially analyzed using GenePix software (Axon Instruments). Additionally, files were loaded onto the Longhorn Array Database (Killion et al. 2003) that normalized each array so that the mean Cy5/Cy3 ratio of median spot intensity was equal to one, and allowed for the compilation of multiple arrays for data retrieval. Log2 ratio of median spot intensity of background-subtracted and normalized data was used as the dependent variable. Poor quality spots detected by GenePix software were removed prior to statistical analysis. Log2 normalized ratios for each gene were averaged across 4–8 slides and compared to zero using a z-statistic. The resulting P-values were used to estimate False Discovery Rate (FDR), an approach to the multiple comparisons problem (Storey 2002). We used a free-access Q-value software to detect differentially expressed transcripts that passed a significance criterion at a specified FDR level. Default program parameters (lambda ranged from 0 to 0.95; smoother p0 method) were employed. Only transcripts with scores that surpassed a threshold of FDR = 0.5 and with z-test P-value < 0.05 were selected for further analysis.

In silico analysis of candidate genes

WebQTL database

WebQTL Database (www.webqtl.org) is a web-based resource for complex trait analysis (Chesler et al. 2003; Wang et al. 2003) that currently contains data for approximately 600 published phenotypes, including data on ethanol-induced CTA. WebQTL also includes estimates of expression for transcripts measured in triplicate arrays across a panel of 35 strains that include C57BL/6J (B), DBA/2J (D), their F1 hybrid and 32 BXD RI strains (Taylor et al. 1999; Williams et al. 2001). These measurements of mRNA expression were obtained in naïve mice using the Affymetrix U74Av2 microarrays (Chesler et al. 2003). Pooled samples from trios of male or female mice included the forebrain (except the olfactory bulb) and the entire midbrain (see http://webqtl.org/dbdoc/U74Av2Mas5_December03.html for sample and preparation details). This brain area contained the same brain regions used in KO experiments and also included some other brain structures. In the present study we have used WebQTL to extract genetic correlations (Pearson’s product moment; for review see Crabbe et al. 1990) between two measures of ethanol-induced CTA and estimates of transcript abundance in the CNS of the same strains of mice. These behavioral data are listed in the Published Phenotype database and originate from a study by Risinger and Cunningham (1998) in which conditioned taste aversion to saccharin solution was paired with injections of either 2 g/kg (CTA-2) or 4 g/kg (CTA-4) ethanol. These authors studied the parental strains and 20 BXD RI strains. We used their saccharin consumption-independent residual scores (record numbers 9756038.19 and 9756038.2 in WebQTL) and refer to these traits as CTA-2 and CTA-4. These traits were correlated in WebQTL with 12422 corresponding mRNA estimates listed in database UTHSC Brain mRNA U74Av2 [Dec03] MAS5. Pearson correlation P-values were used to estimate FDR. Again, transcripts with scores that surpassed a threshold of FDR = 0.5 and with correlation P-value < 0.05 were selected for further analysis. These two CTA scores themselves have a Pearson r of 0.57.

Gene expression meta-analysis

Lists of significantly regulated genes from three genetic models: Girk2 KO, Penk KO and WebQTL-based BXD RI strains were compared. Gene sequence accession numbers from two platforms (cDNA and Affymetrix) were matched by converting the accession numbers to the Locus Link IDs (http://www.ncbi.nlm.nih.gov/LocusLink) using the
web-based Database for Annotation, Visualization and Integrated Discovery (DAVID, http://apps1.niaid.nih.gov/David). Genes regulated in both KO models and correlated significantly with at least one of the two ethanol-induced CTA measures in BXD RI panel were considered potentially relevant in regulation of ethanol-induced CTA.

Over-representation analysis
Sequence accession numbers of significantly regulated genes were submitted to DAVID for functional annotations. DAVID uses Gene Ontology (GO) Consortium Database (http://www.godatabase.org) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Database (http://www.kegg.org) to assign genes to functional groups. A standalone version of web-based Expression Analysis Systematic Explorer (EASE) software version 2 (http://david.niaid.nih.gov/david/ease.htm) was used for GO-based functional group over-representation analysis. The EASE software uses a modification of Fisher Exact Test to compare representation (number of genes) of a certain functional group on a list of significantly regulated genes to representation of this functional group on a total list of all genes used for microarrays. In other words, it estimates whether a certain biological function/pathway is represented by more significantly regulated genes than expected by chance. In addition, a similar over-representation analysis was carried out for KEGG-based functional groups using $\chi^2$ statistics.

Results

Gene expression profiles
Our initial analytic objective was to ensure more complete capture of true positives and to tolerate a comparatively high rate of false positives in each of the three principle data sets. For each experimental group we accepted a false discovery rate of 0.5 and lower. This value estimates an approximate proportion of false positives in lists of genes considered to be significant. The overall results of the microarray experiments show that 734 and 2820 transcripts were affected by the deletion of Girk2 ($P<0.026$; FDR = 0.5) and Penk ($P<0.05$; FDR = 0.2) genes, respectively (Fig. 1). The deletion of functional Girk2 or Penk transcript did not result in dramatic changes in gene expression, as the average fold-change for clones that reached significance was 1.2-fold. Interestingly, a greater number of significantly changed genes were detected in Penk KO mice (Fig. 1b) compared to Girk2 null mutants (Fig.1a). In addition, the number of up-regulated genes was greater than the number of down-regulated genes in both KO lines.

In silico analysis of candidate genes

WebQTL database
Genetic correlations between CTA-2 and CTA-4 measures of ethanol-induced CTA and 12422 mRNA abundance measures extracted from the BXD RI database were calculated. The CTA-4 variable was significantly correlated with 1027 gene transcripts ($P<0.05$; FDR = 0.38). The Q-value software was unable to detect any CTA-2 correlations that would surpass the FDR = 0.5 threshold. One possible reason for this is a comparatively low genetic variability of the CTA-2 measure (Risinger & Cunningham 1998). It appeared that treatment with 4 g/kg ethanol resulted in more diverse CTA responses among the BXD RI strains, which might lead to a better detection of significant associations with transcript abundance. Therefore, transcript correlations with the CTA-2 variable were not included in further analysis. The other three lists of significantly regulated genes (Penk KO, Girk2 KO and CTA-4) were further subjected to gene expression meta-analysis and over-representation analysis.

Gene expression meta-analysis
Table 1 shows eight transcripts that were significantly regulated in both KO lines and were genetically correlated with CTA-4 measure, implying that they are potential candidates for regulation of ethanol-induced CTA. The polarity of association between transcript abundance and magnitude of CTA was similar in all three experimental groups for Penk1 transcript only. Lower gene expression of Penk1 was associated with greater CTA. It is important to note that for some of the

Figure 1: Frequency distributions of significantly regulated genes in two KO lines ($z$-test $P<0.05$; FDR ≤ 0.5). A greater amount of significantly changed genes were detected in Penk KO mice (b) compared to Girk2 null mutants (a). The magnitude of changes in either knock-out model was less than twofold for the majority of genes. Number of up-regulated genes was greater than number of down-regulated genes in either KO line.
detected genes there are multiple probesets in the WebQTL database. For example, there are three probesets for the Penk1 gene. Only the probeset that showed the highest expression values correlated significantly with the CTA-4 variable. The probeset with moderate expression showed a robust trend to correlate, while the set with the lowest expression values, which most likely represented noise, did not correlate with CTA-4 (data not shown). This pattern was also evident for other detected transcripts.

Over-representation analysis

Three separate lists of significantly regulated genes, one from each KO line and one from the BXD RI set (CTA-4) were subjected to over-representation analyses, allowing us to detect functional groups that are potentially relevant to regulation of ethanol-induced CTA. Table 2 shows functional categories that were over-represented in all three genetic models. One of the two functional categories that were detected in all three models studied was KEGG-based Phosphatidylinositol Signaling System pathway, implying that the phosphoinositide signal-transduction cascade is involved in regulation of ethanol-induced CTA. A schematic diagram of this pathway is shown in Fig 2. Each box on this diagram represents a category of proteins with similar function. For example, PTEN combines a number of protein tyrosine phosphatases of different types (for details see www.kegg.org). Most of the protein categories detected in our study were present in only one or two genetic models. Although the PTEN category was represented by significant genes detected in all three genetic models, no protein phosphatase of the same type was detected more than twice. In fact, protein kinase C (PKC) beta isoform was the only gene from this pathway detected in all three genetic models. Interestingly, several other PKC isoforms were detected on one or two of the three gene lists. PKC lambda and theta were regulated in Girk2 and Penk KO lines, respectively; PKC alpha was present in both KO gene lists; and transcripts of PKC isoforms eta and mu were genetically correlated with CTA-4 in BXD RI strains.

Discussion

To detect genes that may underlie ethanol-induced CTA we used a strategy of convergent evidence to investigate three independent genetic models tested for large-scale gene expression and behavior. A novel aspect of this approach was to compare transcriptional profiles of two different null mutant lines with differences in this phenotype. In addition, we used the interactive web-based system for complex trait analysis (WebQTL Project) to detect genes that genetically correlated with measures of CTA.

Table 1: Genes that show significant association with ethanol-induced CTA in all three genetic models: Girk2 KO, Penk KO and BXD RI strains. The level of significance determined by a z-test in KO experiments and by Pearson’s correlations for the BXD RI panel is shown as P-values. The P-values in bold indicate positive association between gene expression and ethanol-induced CTA, i.e. higher gene expression is associated with greater CTA. P-values in italic reflect negative correlation between transcript abundance and behavior. Numbers in parentheses indicate absolute fold change (KO vs. respective control). Gene sequence accession numbers are shown in columns 2–4.

Table 2: Functional groups that show significant association with ethanol-induced CTA in all three genetic models: Girk2 KO, Penk KO and BXD RI strains. Three separate lists of significantly regulated genes were subjected to an overrepresentation analyses (see Materials and methods). Numbers in bold are P-values for each group, which resulted from either EASE analysis or a chi-square test. Numbers in parentheses indicate frequencies (%) of genes that belong to a certain functional group: expected frequencies in the first column and observed frequencies in the other four columns.
It was somewhat surprising to find such small changes in expression in both of the two KO lines studied. GIRK2 represents one of the most abundant GIRK subunits in the brain, and with the observed decrease in GIRK1 expression (Signorini et al. 1997; Torrecilla et al. 2002), one would predict that the overall levels of GIRK channels would be reduced. The loss of GIRK channels in neurons could presumably result in a change in neuronal resting membrane potential (Kuzhikandathil & Oxford 2000) or an alteration in intracellular functioning of G-protein-coupled receptors. Proenkephalin is a major neuropeptide, highly expressed in the brain; however, neither the phenotypes of the knockout mice nor the gene expression studies indicate a major effect of the gene deletion.

Although there are many reports of microarray-based gene expression changes in the brain of different mouse models, there are few studies that examined wide-scale transcriptional profiles of null mutant mice. In agreement with our results, some studies reported small changes in gene expression when KO and wild type mice were compared (Poguet et al. 2003; Tudor et al. 2002). For example, Tudor et al. (2002) presented a thorough examination of gene expression profiles of a mouse model of Rett syndrome. The authors studied nine brain regions and used a representative sample size (3–9 mice per genotype per region; 1 mouse per array) to compare transcriptional profiles of Mecp2 null mutant mice and their controls. The results of these comparisons showed that, despite severe physiological abnormalities in KO mice, which ultimately led to death at 2 months, the mutant brains had few genes with changes in expression of more than 1.5-fold. The authors suggested the possibility that neurons are sensitive to subtle transcriptional changes, and that such changes underlie the phenotype. This idea is consistent with recent data from yeast, showing that...
some small transcriptional changes are amplified dramatically at the level of protein (Ghaemmaghami et al. 2003). One functional pathway and eight genes were selected as strong candidates for regulation of ethanol-induced CTA. Our over-representation analysis determined that Phosphatidylinositol Signaling System pathway was represented by more significantly regulated genes than expected by chance in all three genetic models. The phosphoinositide (PI) signal-transduction cascade represents an important second messenger system that is activated by many neurotransmitters and hormones in the central nervous system. The PI pathway is modulated by both acute and chronic ethanol exposure and is believed to trigger some critical molecular events associated with the development of ethanol tolerance and dependence (for review see Greenbaum 2003). In addition, one study showed that mice deficient for epsilon isoform of PKC, a key enzyme in the PI-signaling cascade, consume less alcohol (Hodge et al. 1999). Because we used data from ethanol-naïve mice in the present study, our results suggest that the hereditary differences in the PI-signaling cascade may result in differential sensitivity to the aversive effects of ethanol.

PKC beta gene was detected as a part of the PI signal-transduction pathway and was also differentially regulated in all three genetic models studied. The calcium- and phospholipid-dependent PKC family of enzymes has been implicated in synaptic plasticity and memory, null mutant mice lacking PKC beta exhibited deficits in both cued and contextual fear conditioning (Weeber et al. 2000). Because acquisition of ethanol-induced CTA is dependent on intact learning mechanisms, PKC beta may play a central role in this process.

Preproenkephalin gene was another strong candidate. The Penk deficient mice lacking functional Penk transcript showed increased ethanol-induced CTA, while up-regulation of Penk1 transcript in Girk2 null mutants was associated with decreased CTA. This result is consistent with one previous report, showing that blocking delta opioid receptors resulted in increased ethanol-induced CTA (Froehlich et al. 1998). It is of interest that both candidate genes discussed have functional links with the PI-signaling cascade. While PKC is a key step in this pathway, enkephalin-binding mu-opioid receptors activate three different effectors of the PI-signaling cascade (Polakiewicz et al. 1998).

It is important to note that direction of changes in transcript abundance does not necessarily imply a similar direction in protein abundance and ultimately in expression of function. Several studies showed that the relationships between mRNA and protein levels are complex and depend on specific gene, genetic system and treatment conditions (for review see Greenbaum et al. 2003). They are significantly affected by post-translational mechanisms that may vary from gene to gene. The role of a specific protein in a higher level behavioral function can also depend on activity of other members of underlying pathways. While some gene transcripts could be positively regulated with their corresponding functions, others could show negative correlation. For example, it is not unusual in KO experiments to see a paradoxical increase in the mRNA signal of the KO gene, presumably an attempt of the neuron to up-regulate the lost function. Another example could be provided from the WebQTL database. A transcript of PKC theta was negatively correlated ($r = -0.85; P = 0.0003$) with PKC cytosolic activity, one of the phenotypes measured in a separate experiment (Wehner et al. 1990). Therefore, we can not expect to observe a unidirectional relationship between mRNA abundance and behavior under all conditions and in all genetic models.

The goal of the present study was not to emphasize the direction of transcriptional changes, but to detect general transcriptional regulation associated with the behavior of interest. The transcript–behavior associations do not necessarily imply cause and effect relationships, but simply indicate that neurons are responding to genetic manipulations by changing the transcript level, which could be of important relevance to otherwise altered behavior. Detecting transcript–behavior associations in all three models greatly increases chances for such associations to be functionally relevant in at least one of the three models.

A potential contribution of genetic background to our results also deserves mentioning. All three genetic models used in the present study were generated on substantially different genetic backgrounds, with C57BL/6J genes being the only source of common genetic material. While genetic variance plays a key role in analysis of BXD RI data, introducing such variability to KO lines can mask the effects of genetic mutation, especially in cases when these effects are of small magnitude. For example, fewer significant changes were detected in Girk2 KO compared to Penk KO line (see Fig. 1), which could, in part, result from introduction of a second genetic background in Girk2 mice. In our study, background variation could to some extent be considered as a negative control, resulting in detection of fewer but probably more stable gene–behavior associations. In addition, studying gene expression in a larger brain area in BXD RI strains compared to KO lines might also lead to detection of only strong transcript–behavior associations, because of a possibility that small transcriptional changes could be masked by variable gene expression in different cell types.

In summary, our results show that deletion of either Girk2 or Penk gene produced diverse but small changes in gene expression. Some of these changes may underlie phenotypic abnormalities in these mutants; in particular, some genes that were significantly regulated in both KO models and the BXD RI set might be involved in regulation of ethanol-induced CTA. Although microarray experiments generate a large number of candidate genes and a potential for false positives, we suggest that these problems can be managed to a certain degree by using in silico analyses to establish convergent validity. This approach allows us to focus on a single functional pathway and a few critical genes for future...
studies. PKC beta and Penk genes were detected from a list of possible candidates by means of different in silico procedures and are consistent with other literature reports. Different strategies including pharmacological intervention, knockout methodology and interference RNA blocking can now be used to study functional relationships between these genes and ethanol-induced CTA. Studying gene expression profiles of null mutant mice should be of use to study genetic mechanisms of behavior, especially when two or more different KO lines demonstrate similar behavioral alterations. Creating a central web-based database containing transcriptional and phenotypic data on different KO models should accelerate this type of research dramatically.

References


Microarrays and genetic models to study behavior


Acknowledgments

Supported by the NIAAA INIA program, grants: AA13520, AA13518 and NIH grant AA 06399. Authors would like to thank Elizabeth Osterndorff-Kahanek for verification sequencing, Dr Dayne Mayfield for his comments on initial drafts of this manuscript and also Jonathan Davies, Kerri Jeffers and other members of the Iyer lab for their assistance with preparing microarray slides.